Genetics: Gene Expression

Genomes contain all the genetic information need by an organism, regardless of the circumstance or environment. Generally, however, even single cell organisms use only a fraction of their genome at any time. That is, cells generally express only a fraction of their genome at any given time. This is especially true among the individual cells that make up multicellular organisms where cell specialization (differentiation) dictates that only certain genes be expressed in certain cells, and only at particular times during the organism's life span.

Genes and non-coding DNA.

In most organisms, genes are composed of discrete segments of DNA. Genes code for all of the proteins and structural RNA molecules used by cells within an organism. The coding segments of genes are often flanked by segments, or elements, that provide regulatory information. Regulatory elements control gene expression -- the process that produces a functional protein or RNA molecule from the encoding DNA. Many steps in the control of gene expression are conserved between prokaryotes and eukaryotes, and between simple and complex eukaryotes, although some features differ among these classes of organisms. Except where noted, the discussion below focuses on complex eukaryotes.

Regulatory DNA. Regulatory DNA includes several types of sequence elements that dictate when and how a gene is expressed. Among the most important of these is the promoter -- the sequences that direct the start of transcription (the process that transcribes DNA into RNA; discussed below). Regulatory DNA elements also determine under what environmental, developmental or physiological circumstances a gene is transcribed. Regulatory elements may

stimulate or repress transcription by directly binding specific transcriptional activator or repressor proteins. They may also stimulate or repress transcription indirectly, by binding proteins that alter the compaction of the DNA and hence the accessibility of the promoters and other elements to binding proteins. Promoters and some regulatory elements are generally within a few hundred base pairs of the first coding segment of a gene. Other regulatory elements, particularly those that act indirectly, can be many thousands of base pairs away from genes.

Chromatin. DNA is virtually never naked -- that is, devoid of proteins -- in cells. Rather, DNA in cells is packaged into chromatin -- a complex of DNA and protein. Minimally, DNA is bound by histones, proteins that form a positively charged core around which 150-200 nucleotides of DNA are wrapped. Each core histone-DNA complex defines a nucleosome. Nucleosomes are separated by short DNA segments or linkers, which are bound by a non-core histone. Many other proteins bind nucleosomal DNA. Some of these regulate the higher order organization of nucleosomal DNA into secondary coils or folded threads, which in turn are organized into large loops, the bases of which are anchored to the nuclear matrix. Chromatin loops may define domains of heterochromatin, which is highly compact and relatively inaccessible to the transcription machinery, or euchromatin, which is relatively open and accessible to the transcription machinery.

Non-coding DNA. The coding segments of genes (exons) are often interspersed with relatively large non-coding segments (introns). Introns are initially transcribed into RNA but are subsequently removed by the process of RNA splicing. The functions of introns are not well understood, but at least some introns contain transcriptional control elements. In addition, some

introns contain sequences that bind the nuclear matrix, and hence may influence whether genes are in heterochromatic or euchromatic chromatin loops.

Complex eukaryotic genomes contain other large DNA segments that do not encode genes. Some of these are important structural components of chromosomes (for example, telomeres and centromeres). Others have no obvious structural or regulatory role, but then our understanding of non-coding DNA sequences is still very incomplete.

Flow of genetic information.

Genes are transcribed into RNA by the process of transcription. Some RNA molecules are functional in and of themselves. Examples include the RNA components of certain enzymes (for example, telomerase), transfer RNA (tRNA), which delivers amino acids to the ribosome for protein synthesis, and ribosomal RNA (rRNA), a structural component of the ribosome. Other RNA molecules are intermediaries in protein synthesis. Such RNAs are processed in the nucleus into messenger RNA (mRNA) molecules. mRNAs are then exported to the cytoplasm, where they bind ribosomes and direct synthesis of the encoded protein by the process of translation.

Genes are expressed when their ultimate products (RNAs or proteins) are produced. Sometimes, genes encoding proteins are considered expressed when they are simply transcribed, but it should be remembered that the proteins, not the transcripts, are the functional products.

Control of gene expression.

Gene expression is controlled at multiple levels, including transcription (initiation and elongation), post-transcriptional processing, RNA stability, RNA export and association with

ribosomes, translation (initiation and elongation) and post-translational processing. Each level tends to be highly regulated and complex. Moreover, each level requires the cooperation of both general and cell type-specific proteins.

Just as phenotypic differences among species are due to differences in the genes encoded by their genomes, differences among cell types within an organism are largely due to differences in which genes within a genome are transcribed. The genomes of multicellular organisms contain genes that are transcribed in all or most cell types, as well as genes whose transcription is confined to specific cells. From the earliest stages of embryogenesis, transcription is confined to only certain segments of the genome, depending on the cell type and stage of development. Thus, cellular differentiation -- the process by which cells acquire and maintain specialized functions -- generally entails the differential activation and repression of gene transcription. Differential gene transcription is generally regulated at two broad levels.

Control by transcriptional activators and repressors. All genes are transcribed by the transcription machinery, which consists of a large protein complex. The basal transcription complex contains the core proteins needed to recognize promoter sequences, unwind the DNA duplex, and initiate, elongate and terminate the primary transcript. This complex also contains proteins that recognize certain types of DNA damage, and can recruit proteins to repair the transcribed DNA strand. The basal transcription complex interacts with a large number of specific transcriptional activators and repressors (transcription factors), regulatory proteins that bind elements outside the immediate promoter region. These regulatory proteins, then, dictate whether or not the basal transcription complex initiates transcription.

Some cells express highly specialized transcription factors, which regulate the expression of genes confined to that cell type or its precursors. For example, muscle cells express specific

factors that control the transcription of genes encoding muscle-specific proteins. Although these transcription factors can stimulate muscle-specific gene transcription in some non-muscle cells (for example, fibroblasts), they cannot activate muscle specific gene transcription in many other cell types. Moreover, some cell-type specific genes are controlled by transcription factors that are expressed by many different cell types. Thus, the presence of specific transcriptional activators and repressors alone is generally insufficient for cell type-specific gene transcription. In addition to the presence of specific transcription factors, the target genes must be in an accessible chromatin state.

Control by chromatin state and epigenetic inheritance. In general, genes located within heterochromatin are inaccessible to the transcriptional machinery, and thus are not expressed despite the presence of specific transcriptional activators. Such genes are said to be silenced in order to distinguish them from unexpressed genes in euchromatin. Unexpressed genes in euchromatin remain accessible to the transcriptional machinery, and thus can readily respond to physiological or environmental signals. Silenced genes, by contrast, cannot respond external signals unless the signal includes one to remodel the chromatin (that is, reset the boundaries of heterochromatin and euchromatin). Whether a DNA segment is heterochromatic or euchromatic is generally determined during embryogenesis. The mechanisms that control the state of chromatin are incompletely understood. They include reversible changes to the DNA, such as methylation of cytosine, as well as reversible changes to chromatin proteins, such as acetylation of histones.

The state of chromatin is an important mechanism for initiating and maintaining differential gene expression in multicellular organisms. In adult organisms, the state of the chromatin is generally stably and faithfully maintained, even though chromatin-associated

proteins are transiently stripped from the DNA during the processes of replication or repair. Thus, once the pattern of chromatin is established in a differentiated cell, it is stably inherited from one cell generation to the next. This form of inheritance is termed epigenetic inheritance, since it does not entail irreversible changes to genomic DNA.

Gene expression and aging.

Age-dependent changes in the expression of specific genes have found in virtually all organisms and tissues that have been tested. These changes have been found to occur at all levels, ranging from the initiation of gene transcription to the post-translational modification of proteins. In many cases, it has been difficult to decipher which changes in gene expression are responsible for aging phenotypes, as opposed to which changes are responses, whether adaptive or maladaptive, to primary age-related changes tissues, cells or molecules.

Despite the many age-related changes in gene expression that have been documented, some general principles and common themes have emerged.

First, the expression pattern of many genes do <u>not</u> change with age. This fact has been most thoroughly established by the use of cDNA microarrays, which can assess the levels of mRNA corresponding to hundreds or thousands of genes in a single experiment. The microarray analyses indicate that -- at most -- only few percent of the genes expressed by a given tissue or cell type show an age-dependent increase or decrease in mRNA level.

Second, in many instances, aged cells or tissues appear to be in a chronically stressed state. The origin of this stress is not clear, but may include exogenous or endogenous oxidative damage or subacute inflammation. Whatever the origin, aged cells and tissues frequently show

changes in gene expression that appear to be an adaptive response to stress. In mouse liver, for example, a transcription factor that activates the expression of genes encoding acute phase proteins increases with age. The acute phase response is invoked when tissues are inflamed or oxidatively stressed. Similarly, some heat shock proteins are modestly but constitutively elevated in aged organisms (for example, Drosophila), tissues (for example, mouse liver), and cells (for example, senescent human fibroblasts). The heat shock response is induced by stresses that cause damaged or misfolded proteins to accumulate. The idea that some age-related changes in gene expression are adaptive is supported microarray analyses of cells and tissues from calorically restricted animals. Caloric restriction extends the life span of many organisms. It also reverses some, but certainly not all, age-related changes in gene expression.

Finally, some aged cells and tissues fail to mount an adequate stress response and hence are hypersensitive to stress-induced damage or death. Perhaps the best example of this is the heat shock response. Heat and other stresses induce high levels of heat shock proteins in young cells and tissues. In aged cells and tissues, by contrast, heat and other stresses fail to induce high levels of heat shock proteins. The heat shock response is due to transcriptional activation of the heat shock genes by a specific transcription factor. Aged tissues and cells apparently contain adequate amounts of the transcription factor. However, for reasons that are not understood, the transcription factor fails to bind the heat shock gene promoter element in aged cells and tissues.

Gene expression and life span.

No single change in gene expression has yet been shown to be responsible for limiting or extending the life span of an organism (or even a tissue, for that matter). However, manipulation of general regulators of gene expression has been shown to alter life span, at least in model

organisms. Perhaps the best example of this is the SIR2 protein, which is conserved from yeast to mammals and is responsible for the heterochromatic silencing of genes. In yeast, inactivation of the SIR2 gene results in defective gene silencing, increased illegitimate recombination at the repetitive loci that encode rRNAs, and a shortening of life span. Conversely, an additional copy of the yeast SIR2 gene, which presumably results in more efficient silencing and suppression of recombination, extends yeast life span. Of greater relevance to complex eukaryotes, an additional copy of the SIR2 gene in the model organism *Caenorhabditis elegans*, a multicellular nematode, also extends life span. SIR2 requires NAD (nicotinamide adenine dinucleotide) as a cofactor. Thus, SIR2 may act to coordinate energy utilization, recombination and control of gene expression, all of which may strongly influence organismal life span.

WORD COUNT INCLUDING HEADINGS: 2068

Bibliography

Finkel, Toren, and Holbrook, Nikki J. "Oxidants, Oxidative Stress and the Biology of Ageing." *Nature* 408 (2000): 239-247.

Guarente, Leonard. "Sir2 Links Chromatin Silencing, Metabolism and Aging." *Genes and Development* 14 (2000): 1021-1026.

Han, Eun, Hilsenbeck, Susan G., Richardson, Arlan, and Nelson, James F. "cDNA Expression Arrays Reveal Incomplete Reversal of Age-Related Changes in Gene Expression by Calorie Restriction." *Mechanisms of Ageing and Development* 115 (2000): 157-174.

Hedari, Ahmed R., You, Shenghong, Takahashi, Ryoya, Gutsmann-Conrad, Astrid, Sarge, Kevin D., and Richardson, Arlan. "Age-Related Alterations in the Activation of Heat Shock Transcription Factor 1 in Rat Hepatocytes." *Experimental Cell Research* 256 (2000): 83-93.

Lee, Cheol-Koo, Klopp, Roger C., Weindruch, Richard, and Prolla, Tomas A. "Gene Expression Profile of Aging and Its Retardation by Caloric Restriction." *Science* 285 (1999): 1390-1393.

Papaconstantinou, John. "Unifying Model of the Programmed and Stochastic Theories fo Aging. The Stress Response Genes, Signal Transduction-Redox Pathways and Aging." *Annals of the New York Academy of Sciences* 719 (1994): 195-211.

Rattan, Suresh I. "Synthesis, Modifications and Turnover of Proteins During Aging." *Experimental Gerontology* 31 (1996): 33-47.

Roy, Arun K. "Transcription Factors and Aging." *Molecular Medicine* 3 (1997): 496-504.

Vijg, Jan. "Profiling Aging by Gene Arrays." *Mechanisms in Aging and Development* 112 (1999): 1-4.

Villeponteau, Bryant. "The Heterochromatin Loss Model of Aging." *Experimental Gerontology* 32 (1997): 383-394.